Wtloss notebook

Created by Neil Yetz & Gemma Wallace

Contents

Module 6 Lab Activity for PSY 652

Load Libraries

```
library(ppcor)
library(psych)
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
## Warning: package 'tibble' was built under R version 3.6.3
## Warning: package 'tidyr' was built under R version 3.6.3
## Warning: package 'dplyr' was built under R version 3.6.3
## Warning: package 'forcats' was built under R version 3.6.3
```

```
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 3.6.3
```

```
library(apaTables)
```

Import data

```
wtlossa <- read_csv("wtloss_parta.csv")
wtlossb <- read_csv("wtloss_partb.csv")
```

Merge data

```
wtloss<-merge(x=wtlossa, y=wtlossb,by="id")
```

Describe Data

Via base R's summary function

```
summary(wtloss)
```

```
##        id            lbslost            caldef           selfeff
##  Min.   :  1.00   Min.   :-3.190   Min.   :-2.266   Min.   :-3.01976
```

```
##   1st Qu.: 25.75   1st Qu.: 1.623   1st Qu.: 7.252   1st Qu.:-0.52594
##   Median : 50.50   Median : 3.140   Median :10.690   Median : 0.11935
##   Mean   : 50.50   Mean   : 3.063   Mean   :10.804   Mean   : 0.06383
##   3rd Qu.: 75.25   3rd Qu.: 4.697   3rd Qu.:13.893   3rd Qu.: 0.71565
##   Max.   :100.00   Max.   : 7.565   Max.   :23.572   Max.   : 2.71836
```

Via psych's describe function

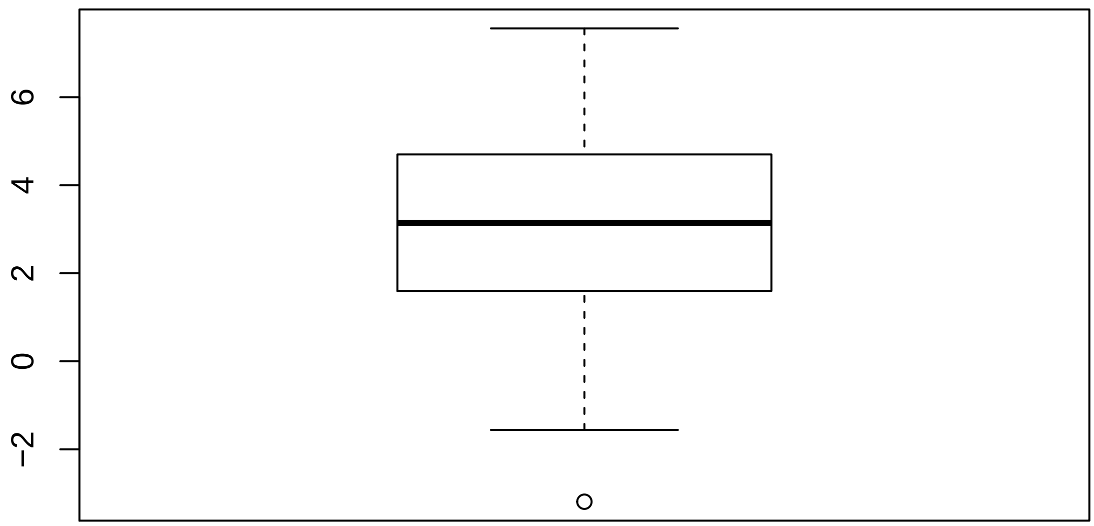
```
describe(wtloss)
```

```
##         vars   n  mean    sd median trimmed   mad   min    max range  skew
## id         1 100 50.50 29.01  50.50   50.50 37.06  1.00 100.00 99.00  0.00
## lbslost    2 100  3.06  2.10   3.14    3.12  2.31 -3.19   7.56 10.76 -0.32
## caldef     3 100 10.80  5.14  10.69   10.86  4.90 -2.27  23.57 25.84 -0.07
## selfeff    4 100  0.06  1.11   0.12    0.07  0.95 -3.02   2.72  5.74 -0.16
##         kurtosis   se
## id         -1.24 2.90
## lbslost    -0.32 0.21
## caldef      0.01 0.51
## selfeff     0.33 0.11
```

Look for outliers with boxplots

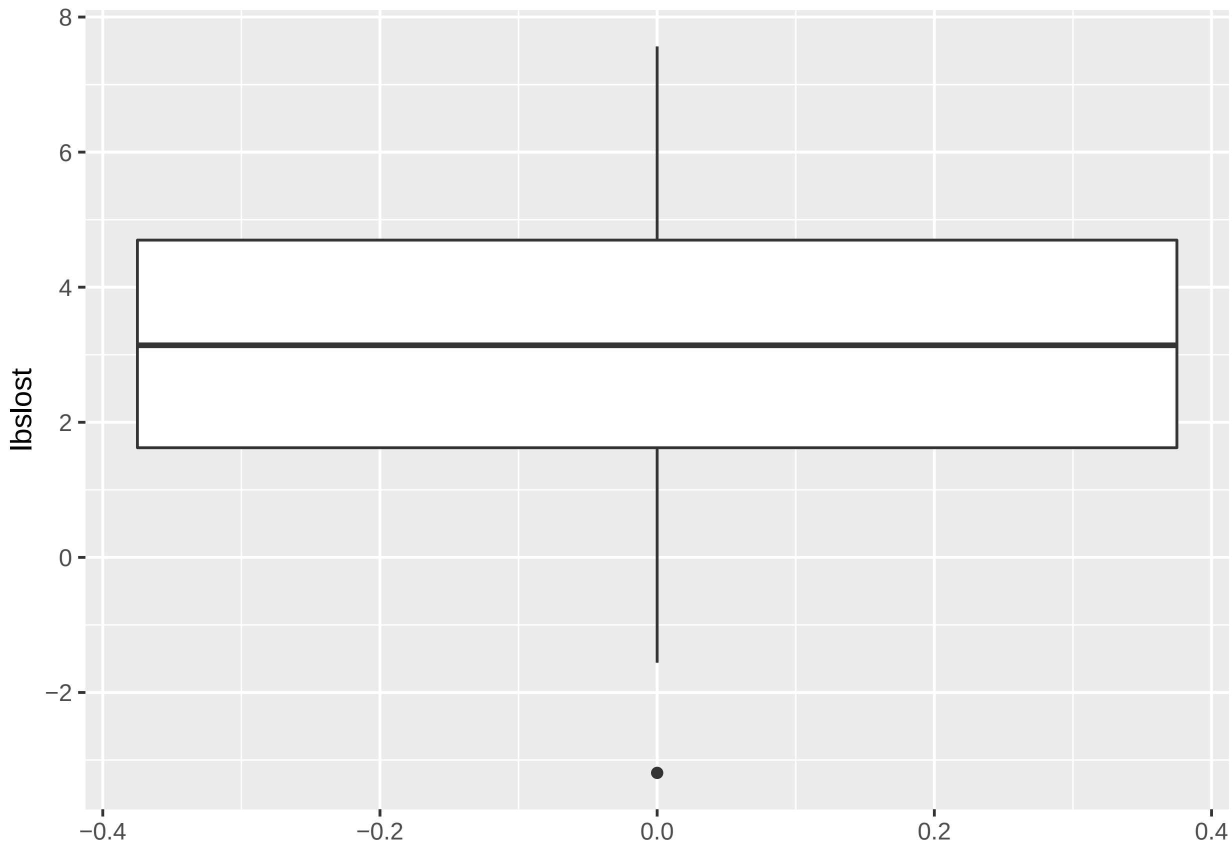
Via Base R's boxplot function

```
boxplot(wtloss$lbslost)
```



Via ggplot2's ggplot function

```
ggplot(wtloss, aes(y = lbslost)) +
  geom_boxplot()
```

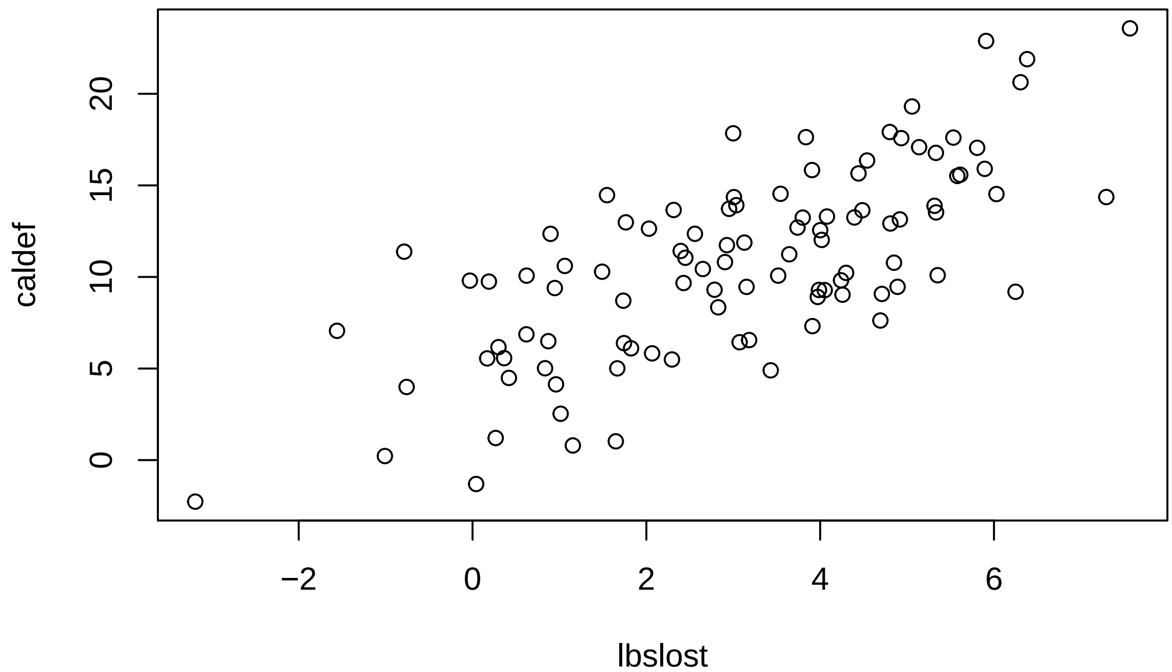


The boxplot indicates that there is one outlier for lbslost, with a value of -4. This indicates that one participant gained 4 pounds during the course of this study.

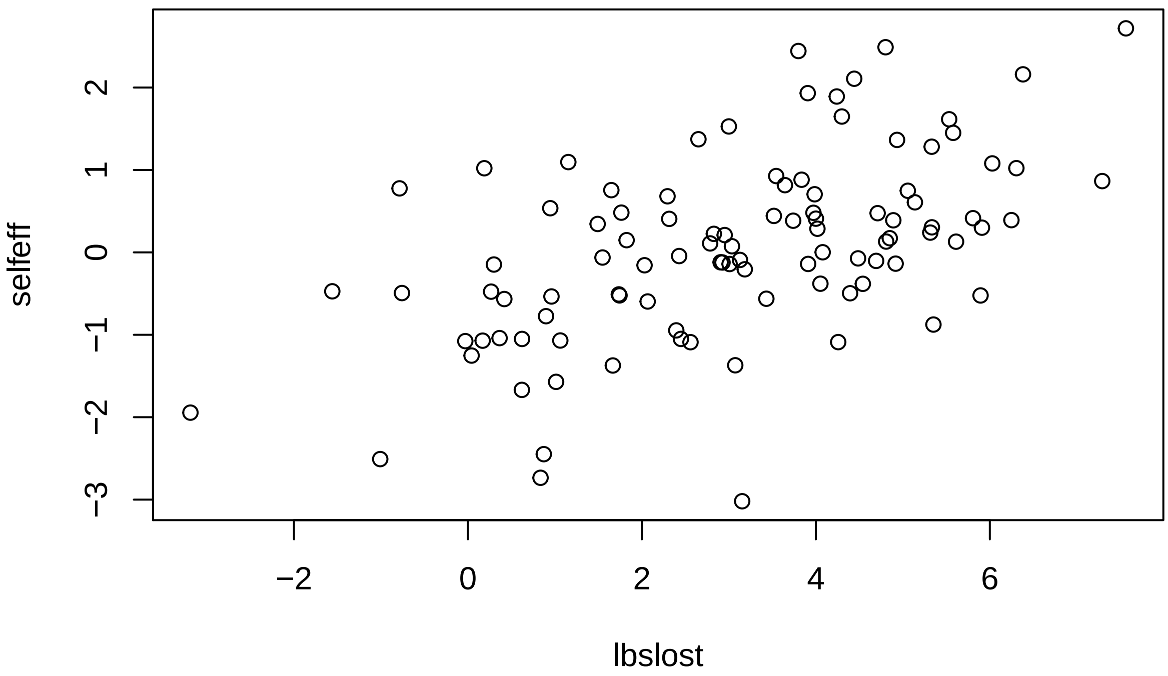
Plot relationships between variables

Via Base R's plot function

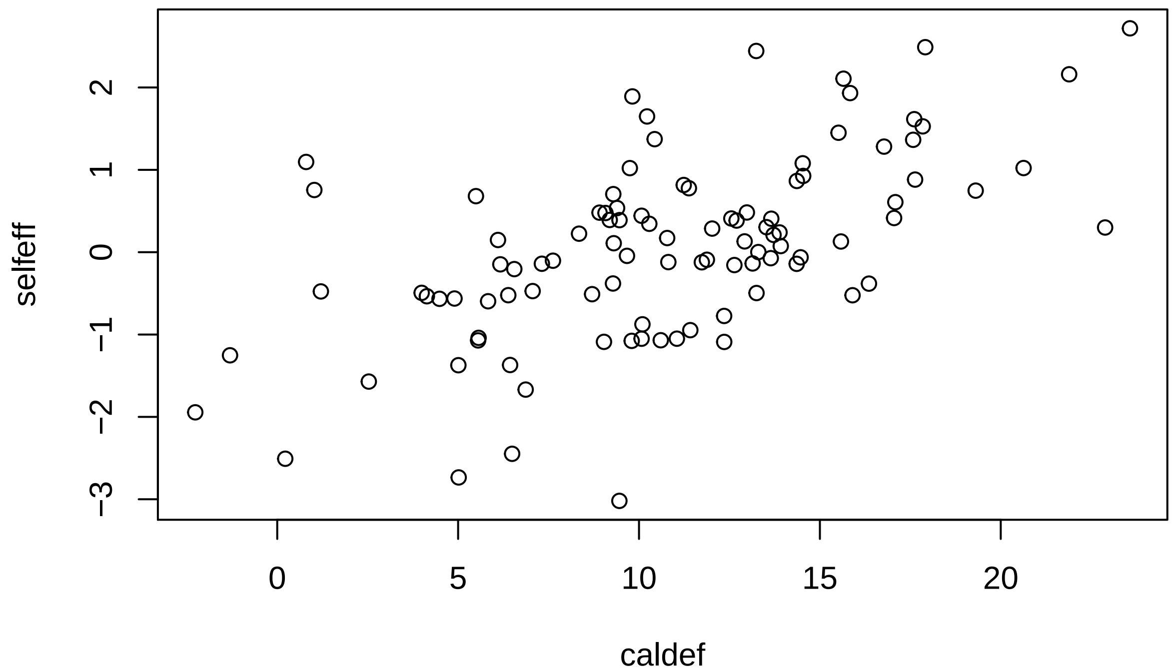
```
attach(wtloss)
plot(lbslost, caldef )
```




```
plot(lbslost, selfeff)
```




```
plot(caldef , selfeff)
```

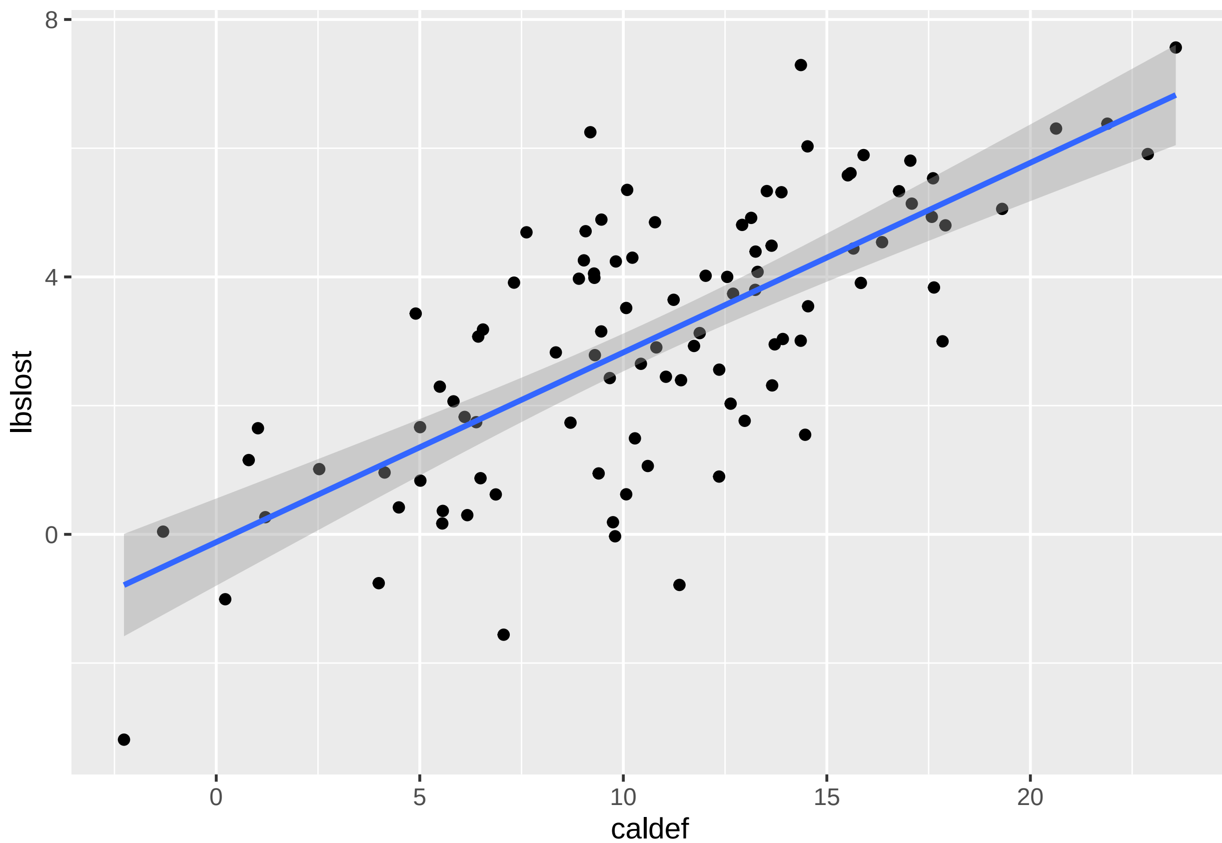



```r
detach(wtloss)
```

Via ggplot2's ggplot function

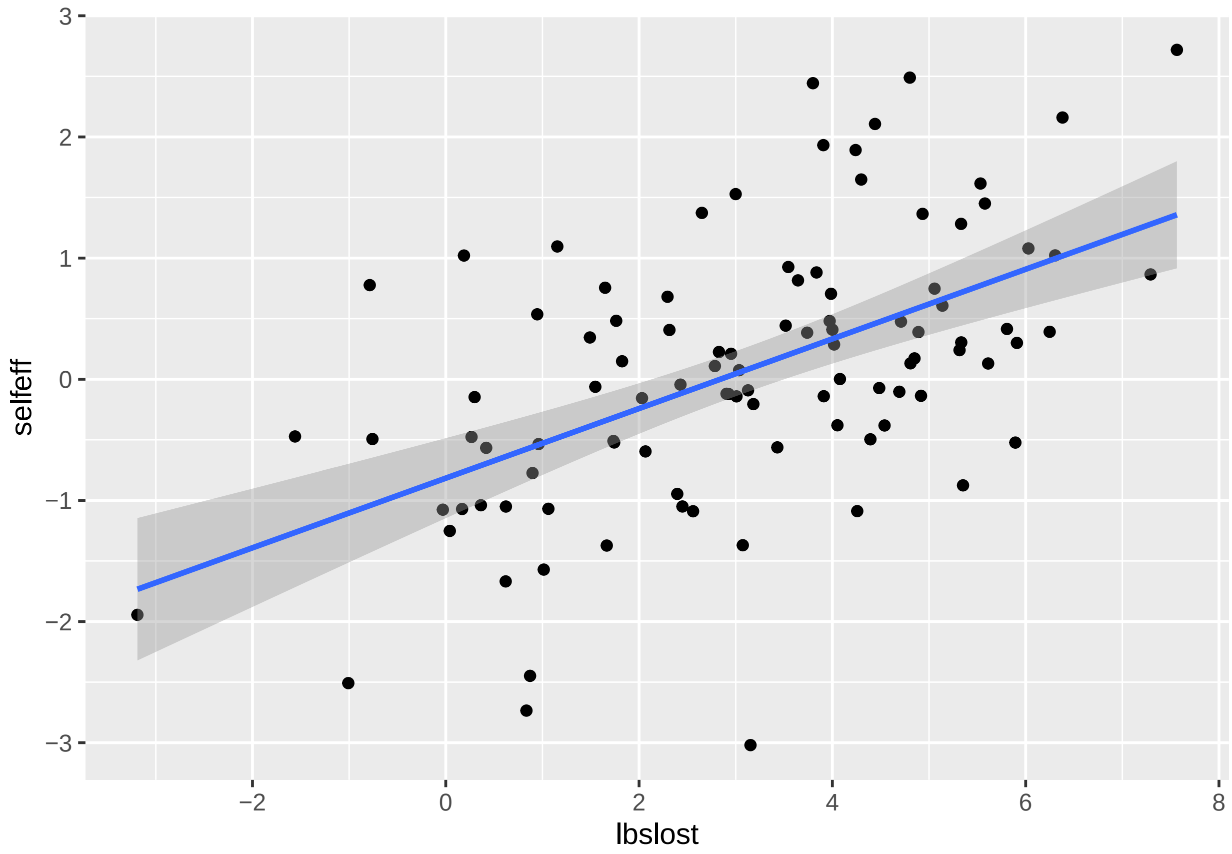
```r
ggplot(wtloss, aes(x = caldef, y = lbslost)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



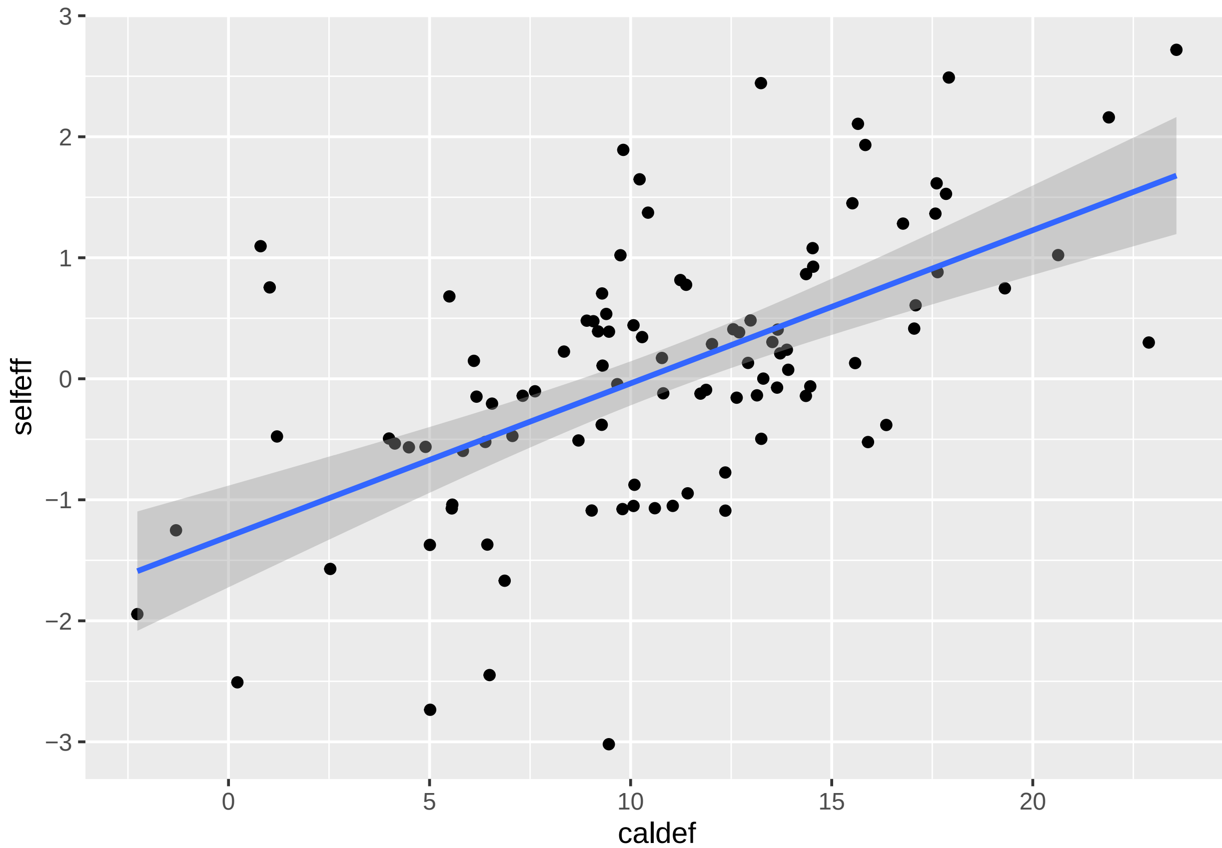

```r
ggplot(wtloss, aes(x = lbslost, y = selfeff)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(wtloss, aes(x = caldef, y = selfeff)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Correlations

Pearson's Correlations Matrices

Via Base R's cor function

```
cor(wtloss)
```

```
##                  id    lbslost      caldef     selfeff
## id       1.00000000 -0.1425016 -0.07237732 -0.05700845
## lbslost -0.14250157  1.0000000  0.72174587  0.54278360
## caldef  -0.07237732  0.7217459  1.00000000  0.58508654
## selfeff -0.05700845  0.5427836  0.58508654  1.00000000
```

Via apaTables' apa.cor.table function

```
apa.cor.table(wtloss, "wtloss correlations.doc", show.conf.interval = TRUE)
```

```
##
##
## Means, standard deviations, and correlations with confidence intervals
##
##
##   Variable    M     SD    1           2           3
##   1. id       50.50 29.01
##
```

```
##   2. lbslost 3.06  2.10  -.14
##                          [-.33, .06]
##
##   3. caldef  10.80 5.14  -.07        .72**
##                          [-.27, .13] [.61, .80]
##
##   4. selfeff 0.06  1.11  -.06        .54**      .59**
##                          [-.25, .14] [.39, .67] [.44, .70]
##
##
## Note. M and SD are used to represent mean and standard deviation, respectively.
## Values in square brackets indicate the 95% confidence interval.
## The confidence interval is a plausible range of population correlations
## that could have caused the sample correlation (Cumming, 2014).
## * indicates p < .05. ** indicates p < .01.
##
```

Correlation coefficients measure the direction and strength of the tendnecy for two variables to vary together. Correlation coefficients do not apply causation.

The correlation coefficient between lbslos and caldef is .72, and this is significant at $p<0.01$. Thus, there is a strong, positive (uphill) correlation between lbslost and caldef.

The correlation coefficient between lbslost and selfeff is .54, and this is significant at $p<0.01$. Thus, there is a strong, positive (uphill) correlation between lbslost and selfeff.

The correlation between caldef and selfeff is .59, and this is significant at $p<0.01$. Thus, there is a strong, positive (uphill) correlation between caldef and selfeff.

The correlation between lbslost and caldef is larger than the correlations between lbslost and selfeff and between caldef and selfeff.

Get partial correlation

```r
attach(wtloss)
pcor.test(lbslost, caldef, selfeff)
```

```
##    estimate      p.value statistic   n gp  Method
## 1 0.5933978 9.634079e-11  7.260806 100  1 pearson
```

```r
detach(wtloss)
```

Partial correlation first removes from both the outcome (lbslost) and predictor (caldef) all variance which may be accounted for by a third variable (in this case selfeff). Then it correlates the remaining variance of caldef with the remaining variance of lbslost. Here, the partial correlation between lbslost and caldef is .593 and this is significant at $p<0.01$. Thus, there is still a strong, positive (uphill) correlation between lbslost and caldef when controlling for or holding constant the effects of selfeff. This correlation is smaller than when the effects of selfeff were not removed (.593 compared to 0.722).

Get semipartial correlation

```r
attach(wtloss)
spcor.test(lbslost, caldef, selfeff)
```

```
##    estimate      p.value statistic   n gp  Method
## 1 0.4983786 1.525307e-07  5.661694 100  1 pearson
```

```r
detach(wtloss)
```

Semi-partial correlation first removes from the predictor (caldef) all variance which may be accounted for by the other predictors (in this case selfeff). Then it correlates the remaining variance of the predictor (caldef) with y (lbslost). Here the semi-partial correlation between caldef and lbslost is .498, and this is significant at $p<0.01$. Thus, there is still a positive (uphill) relationship between lbs and caldef, but this correlation is smaller (now moderately strong; 0.498 compared to 0.722) than when the effects of selfeff on caldef were not controlled for.

Note: the semi-partial correlation is often considered to be a better indicator of the "actual relevance" of a predictor, because it is scaled relative the total variability of the outcome variable. The squared semi-partial correlation will give you the proportion of unique variance accounted for by the predictor, relative to the total variance of Y. This is a key concept in Multiple Linear Regression models.

Fit a Simple Linear Regression (SLR) Model

```r
mod1 <- lm(lbslost ~ caldef, data = wtloss)
```

Display SLR model results

Via base R's summary function

```r
summary(mod1)
```

```
##
## Call:
## lm(formula = lbslost ~ caldef, data = wtloss)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0197 -0.9480  0.0838  1.1227  3.6602
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.12079    0.34127  -0.354    0.724
## caldef       0.29473    0.02855  10.323   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.46 on 98 degrees of freedom
## Multiple R-squared:  0.5209, Adjusted R-squared:  0.516
## F-statistic: 106.6 on 1 and 98 DF,  p-value: < 2.2e-16
```

Via olsrr's ols_regress function

```r
ols_regress(mod1)
```

```
##                         Model Summary
## ----------------------------------------------------------------
## R                         0.722       RMSE              1.460
## R-Squared                 0.521       Coef. Var        47.647
## Adj. R-Squared            0.516       MSE               2.131
## Pred R-Squared            0.504       MAE               1.149
```

```
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##               Sum of
##              Squares        DF    Mean Square       F         Sig.
## ---------------------------------------------------------------------
## Regression    227.034         1        227.034    106.558    0.0000
## Residual      208.801        98          2.131
## Total         435.835        99
## ---------------------------------------------------------------------
##
##                                   Parameter Estimates
## ---------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower    upper
## ---------------------------------------------------------------------------------------
## (Intercept)    -0.121         0.341                 -0.354    0.724    -0.798    0.556
##      caldef     0.295         0.029        0.722    10.323    0.000     0.238    0.351
## ---------------------------------------------------------------------------------------
```

The intercept represents the value of y when x = 0. Here, the intercept for caldef is -0.121, indicating that lbslost = -0.121 when caldef = 0.

The beta for a predictor represents the expected change in y for a 1-unit increase in x. Here, the beta for caldef is 0.295, indicating that lbslost is expected to increase by 0.295 for every 1-unit increase in caldef. The standard error for this estimate is 0.029, indicating that the average distance that the observed values fall from the predicted regression line is 0.029. The upper and lower confidence intervals for this value indicate that there is a 95% liklihood that the true value for this beta estimate falls between 0.238 and 0.351. Because the 95% confidence interval for the estimate does not include zero, and becuase $p < .001$, this effect is significant.

The model R^2 value represents the proportion of variability in y that the model explains. Here, the model R^2 is 0.521, indicating that the predictor (caldef) explains 52.1% of the variability in lbslost.